A no fee

07 621092

GI5181

501

METHOD OF INCREASING YIELD OF MATURE PROTEINS IN MAMMALIAN CELLS

This invention relates generally to an improved method of producing mature proteins in mammalian cells, and more specifically, to a method of enhancing or increasing the extent of functional polypeptides, thereby increasing yields of mature biologically active proteins.

Background of the Invention

Many eukaryotic proteins are naturally synthesized as larger precursor polypeptides, requiring further specific proteolytic processing for full maturation prior to secretion. In many cases, this processing is also essential for full biological activity of the mature protein. Cleavage of these precursors frequently occurs at sites marked by paired basic amino

CERTIFICATE UNDER 37 CFR 1.10

I hereby certify that this paper is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" Service under 37 CFR 1.10 on the date indicated below and is addressed to Hon. Commissioner of Patents and Trademarks, Washington, DC 20231.

Signature Mary E. Bak
Date November 26, 1990
Express Mail No. OB09278293

acid residues, e.g. Lys-Arg and Arg-Arg. [Dickerson et al, J. Biol. Chem., 265:2462 (1990); Achsletter et al, EMBO J., 4:173 (1985); Mizuno et al, Biochem. Biophys. Res. Commun., 144:807 (1987)].

Cleavage at the site of a paired basic amino acid sequence removes many propeptides which function in a variety of roles in the processing of the mature protein. In certain cases the propeptide can mediate correct folding and disulfide bond formation within the protein sequence. In other cases the presence of the propeptide appears to be involved in γ-carboxylation of glutamic acid residues in vitamin K-dependent coagulation factors. γ-carboxylated proteins include Factor IX and Protein C, and certain bone-specific proteins, such as bone Gla protein/osteocalcin. The propeptide can also direct intracellular targeting and regulate the coordinate synthesis of multiple mature peptides from a single precursor polypeptide.

The sequences of the propeptide domains of certain vitamin K-dependent blood coagulation proteins have been published [See, Furie et al, Cell, 53:505 (1988)] and the size of the propeptide has been established for both Factor IX and Protein C.

Factor IX is a zymogen of a serine protease that is an important component of the intrinsic pathway of the blood coagulation cascade. The protein is synthesized in the liver and undergoes extensive co- and post-translational modification prior to secretion. These modifications involve endoproteolytic processing to remove the pre- and pro-peptides, glycosylation, vitamin K-dependent γ-carboxylation of 12 amino-terminal glutamic acid residues and β-hydroxylation of a single aspartic acid residue.

The γ-carboxyglutamic acid residues confer metal binding properties on the mature Factor IX protein and may function similarly in the processing of the other vitamin K-dependent blood clotting proteins. These γ-carboxyglutamic acid residues are essential for coagulant activity. The gamma-carboxyglutamate (GLA) domain of Factor IX has also been identified as a major requirement for cell binding [Derian et al, J. Biol. Chem., 264(12):6615-6618 (1989)].

With the advance of genetic engineering, many eukaryotic proteins are being produced recombinantly in selected cell lines, particularly mammalian cell lines. For example, Chinese Hamster Ovary (CHO) DUKX cell lines producing recombinant Factor IX at high antigen levels

(20 μg/ml/day) have been isolated. However, only 1-2% of that recombinant protein is γ-carboxylated, and therefore biologically active, in the presence of vitamin K3 [Kaufman et al, J. Biol. Chem., 261(21):9622-28 (1986)]. Additionally, amino-terminal sequencing of the recombinant protein has found that 50% of the recombinant Factor IX produced by the CHO cells retain the propeptide [Derian et al, J. Biol. Chem., 264(12): 6615-18 (1989)]. Presumably, the endoproteolytic processing enzyme of the CHO cells directing this cleavage was either saturated or simply inefficient in its function.

Despite the fact that several processing enzymes have been proposed as being involved in the propeptide processing reactions, the enzyme or enzymes responsible for these endoproteolytic cleavages in mammalian cells have not been fully characterized.

The purification of proprotein cleavage enzymes has been hampered by their low levels of activity in mammalian tissue and by their membrane-associated nature. Purification of these specific proteases has been complicated additionally by non-specific cleavage of the assay substrates in vitro, and by contaminating proteases such as those released from lysosomes.

The yeast enzyme Kex2 is a membrane-bound, Ca^{++}-dependent serine protease which functions late in the secretory pathway of Saccharomyces cerevisiae, cleaving the polypeptide chains of prepro-killer toxin and prepro-α-factor at the paired basic amino acid sequences of Lys-Arg and Arg-Arg. [Julius et al, Cell, 37:1075 (1984); Julius et al, Cell, 36:309 (1984)].

When expressed in mammalian cells, yeast Kex2 endopeptidase reportedly cleaved a neuroendocrine prohormone [Thomas et al, Science, 241:226-230 (1988)]. Foster et al, Thrombosis and Haemostasis, 62:321 (1989) have reported that the yeast gene product of Kex2 cleaves the Protein C precursor to a 2-chain form when the yeast endoprotease of the Kex2 gene and the wild-type Protein C precursor are coexpressed. However, propeptide processing and the effect of Kex2 expression have not been studied.

Recently, a human insulinoma cDNA encoding a mammalian subtilisin-like protease, designated PC2, has been implicated in the endoproteolytic processing of prohormones based on its homology to the yeast Kex2 protease [Smeekens et al, J. Biol. Chem., 265:2997 (1990)]. To date, however, no functional activity has been demonstrated for the PC2 clone.

The availability of the complete Kex2 gene sequence also allowed the detection of significant homology between the Kex2 protein and "furin", the product of the partially characterized human fur gene, a gene in the immediate upstream region of the c-fes/fps proto-oncogene [Roebroek et al, EMBO J., 5:2197 (1986)]. The complete nucleotide sequence of the putative coding region of the fur gene has been reported. Upon comparison, the human fur gene product has demonstrated structural homology with the subtilisin-type serine protease encoded by the Kex2 gene of the yeast S. cerevisiae [van den Ouweland et al, Nucl. Acids Res., 18(3):664 (1990)]. However, no evidence of the expression of fur was reported.

There remains a need in the art for a method of increasing the efficiency of proteolytic processing of precursor polypeptides in mammalian cells.

Summary of the Invention

In one aspect, the present invention provides a method for increasing the efficiency of, or otherwise enhancing the production of, a functional, mature protein, which protein requires processing of a pro-peptide form for biological activity. The method may be

used for the production of γ-carboxylated proteins. The invention may also be used for the processing of other proteins, not requiring gamma carboxylation, leading to higher levels of biologically active or otherwise useful proteins.

The method may be performed by transfection into a selected host cell line of one or more expression vectors containing a paired basic amino acid cleaving enzyme (PACE) DNA sequence (SEQ ID NO: 1) and a DNA sequence encoding the selected proprotein, each sequence operably linked to a heterologous expression control sequence, or by transfection of the PACE DNA (SEQ ID NO: 1) into a host cell line known to express the desired protein or by transfection of a DNA for the desired protein into a cell known to express PACE (SEQ ID NO: 2).

Of use in the present invention is a recombinant DNA molecule comprising a DNA sequence encoding PACE (SEQ ID NO: 1) or a homolog thereof. The DNA molecule provides the PACE DNA (SEQ ID NO 1) in operative association with a regulatory sequence capable of directing the replication and expression of PACE (SEQ ID NO: 2) in a selected host cell.

Another aspect of the invention includes a recombinant DNA molecule comprising a DNA sequence encoding PACE (SEQ ID NO: 1) and a DNA sequence encoding

a selected proprotein requiring complete processing for biological activity, both DNA sequences being in operative association with one or more heterologous regulatory sequences capable of directing the replication and expression of PACE (SEQ ID NO: 2) and the selected propeptide in a selected host cell. In one embodiment the selected proprotein requires γ-carboxylation for activity.

In a further aspect the present invention provides a host cell containing and capable of expressing DNA sequences encoding PACE (SEQ ID NO: 1) and a selected precursor polypeptide which is capable of producing high levels of active, mature protein. The cell line may be transfected with the recombinant DNA molecule(s) described above. This cell line may be cultured under appropriate conditions permitting expression of the recombinant DNA. The expressed selected protein is then harvested from the host cell or culture medium by suitable conventional means. This claimed process may employ a number of known eukaryotic, preferably mammalian cells, as host cells for expression of the protein.

Other aspects and advantages of this invention are apparent from the following detailed description of the invention.

Detailed Description of the Invention

PACE, an acronym for paired basic amino acid cleaving enzyme, is a propeptide-cleaving enzyme originally isolated from a human liver cell line. A DNA sequence encoding PACE (SEQ ID NO: 1) (or furin) was published in A.M.W. van den Ouweland et al, Nucl. Acids Res., 18(3):664 (1990), and is reported below in Table I.

It should be understood that the enzyme PACE (SEQ ID NO: 2) as described herein may be encoded by DNA sequences that differ in sequence from this published sequence (SEQ ID NO: 1) due to natural allelic variations or synthetically produced modifications. Provided that the biological activities of mediating propeptide cleavage and γ-carboxylation are retained in whole or part despite such modifications, this invention encompasses the use of all such DNA sequences. The term "PACE" as used herein thus encompasses the peptide and DNA sequences specifically disclosed herein as well as analogs thereof retaining PACE biological activity.

Expression of PACE (SEQ ID NO: 2) in host cells can improve the efficiency of cleavage of a proprotein between the dibasic residues Lys-Arg, Lys-Lys or Arg-Arg into its mature form, resulting in high level expression of the mature protein. Host cells for this expression include preferably mammalian cells for expression of

mammalian proteins. The inventors have now surprisingly discovered that co-expression of PACE (SEQ ID NO: 2) with proteins requiring γ-carboxylation for biological activity permits the expression of increased yields of functional, biologically active mature proteins in eukaryotic, preferably mammalian, cells. The establishment of cell lines which express PACE (SEQ ID NO: 2) provides a convenient and efficient mechanism for the high level production of more completely processed and biologically active proteins.

Table I

Published coding sequence of PACE (furin) (SEQ ID NO: 1)/

(SEQ ID NO: 2)

```
ATG GAG CTC AGG CCC TGG TTC CTA TGG GTC GTA CCA CCA   39
Met Glu Leu Arg Pro Trp Leu Leu Trp Val Val Ala Ala
                5                  10

ACA GGA ACC TTG GTC CTG CTA GCA GCT GAT GCT CAG GGC   78
Thr Gly Thr Leu Val Leu Leu Ala Ala Asp Ala Gln Gly
     15                 20                  25

CAG AAG GTC TTC ACC AAC ACG TGG GCT GTG CGC ATC CCT  117
Gln Lys Val Phe Thr Asn Thr Trp Ala Val Arg Ile Pro
            30                  35

GGA GGC CCA GCG GTG GCC AAC AGT GTG GCA CGG AAG CAT  156
Gly Gly Pro Ala Val Ala Asn Ser Val Ala Arg Lys His
 40                  45                  50

GGG TTC CTC AAC CTG GGC CAG ATC TTC GGG GAC TAT TAC  195
Gly Phe Leu Asn Leu Gly Gln Ile Phe Gly Asp Tyr Tyr
          55                  60                  65

CAC TTC TGG CAT CGA GGA GTG ACG AAG CGG TCC CTG TCG  234
His Phe Trp His Arg Gly Val Thr Lys Arg Ser Leu Ser
                70                  75

CCT CAC CGC CCG CGG CAC AGC CGG CTG CAG AGG GAG CCT  273
Pro His Arg Pro Arg His Ser Arg Leu Gln Arg Glu Pro
     80                  85                  90

CAA GTA CAG TGG CTG GAA CAG CAG GTG GCA AAG CGA CGG  312
Gln Val Gln Trp Leu Glu Gln Gln Val Ala Lys Arg Arg
            95                  100

ACT AAA CGG GAC GTG TAC CAG GAG CCC ACA GAC CCC AAG  351
Thr Lys Arg Asp Val Tyr Gln Glu Pro Thr Asp Pro Lys
105                 110                 115

TTT CCT CAG CAG TGG TAC CTG TCT GGT GTC ACT CAG CGG  390
Phe Pro Gln Gln Trp Tyr Leu Ser Gly Val Thr Gln Arg
          120                 125                 130

GAC CTG AAT GTG AAG GCG GCC TGG GCG CAG GGC TAC ACA  429
Asp Leu Asn Val Lys Ala Ala Trp Ala Gln Gly Tyr Thr
                135                 140
```

```
GGG CAC GGC ATT CTG GTC TCC ATT CTG GAC GAT GGC ATC  468
Gly His Gly Ile Val Val Ser Ile Leu Asp Asp Gly Ile
    145                 150                 155

GAG AAG AAC CAC CCC GAC TTG GCA GGC AAT TAT GAT CCT  507
Glu Lys Asn His Pro Asp Leu Ala Gly Asn Tyr Asp Pro
            160                 165

GGG GCC AGT TTT CAT GTC AAT GAC CAG GAC CCT GAC CCC  546
Gly Ala Ser Phe Asp Val Asn Asp Gln Asp Pro Asp Pro
170                 175                 180

CAG CCT CGG TAC ACA CAG ATG AAT GAC AAC AGG CAC GGC  585
Gln Pro Arg Tyr Thr Gln Met Asn Asp Asn Arg His Gly
        185                 190                 195

ACA CGG TGT GCG GGG GAA GTG GCT GCC GTG GCC AAC AAC  624
Thr Arg Cys Ala Gly Glu Val Ala Ala Val Ala Asn Asn
                200                 205

CGT GTC TGT GGT GTA GGT GTG GCC TAC AAC GCC CGC ATT  663
Gly Val Cys Gly Val Gly Val Ala Tyr Asn Ala Arg Ile
    210                 215                 220

GGA GGG GTC CGC ATG CTG GAT GGC GAG GTG ACA GAT GCA  702
Gly Gly Val Arg Met Leu Asp Gly Glu Val Thr Asp Ala
            225                 230

GTG GAG GCA CGC TCG CTG GGC CTG AAC CCC AAC CAC ATC  741
Val Glu Ala Arg Ser Leu Gly Leu Asn Pro Asn His Ile
235                 240                 245

CAC ATC TAC AGT GCC AGC TGG GGC CCC GAG GAT GAC GGC  780
His Ile Tyr Ser Ala Ser Trp Gly Pro Glu Asp Asp Gly
        250                 255                 260

AAG ACA GTG GAT GGG CCA GCC CGG CTC GCC GAG GAG GCC  819
Lys Thr Val Asp Gly Pro Ala Arg Leu Ala Glu Glu Ala
                265                 270

TTC TTC CGT GGG CTT AGC CAG GGC CGA GGG GGG CTG GGC  858
Phe Phe Arg Gly Val Ser Gln Gly Arg Gly Gly Leu Gly
    275                 280                 285

TCC ATC TTT GTC TGG GCC TCG GGG AAC GGG GGG CGG GAA  897
Ser Ile Phe Val Trp Ala Ser Gly Asn Gly Gly Arg Glu
            290                 295

CAT GAC AGC TGC AAC TGC GAC GGC TAC ACC AAC AGT ATC  936
His Asp Ser Cys Asn Cys Asp Gly Tyr Thr Asn Ser Ile
300                 305                 310
```

```
TAC ACG CTG TCC ATC AGC AGC GCC ACG CAG TTT GGC AAC  975
Tyr Thr Leu Ser Ile Ser Ser Ala Thr Gln Phe Gly Asn
        315                 320                 325

GTG CCG TGG TAC AGC GAG GCC TGC TCG TCC ACA CTG GCC 1014
Val Pro Trp Tyr Ser Glu Ala Cys Ser Ser Thr Leu Ala
                330                 335

ACG ACC TAC AGC AGT GGC AAC CAG AAT GAG AAG CAG ATC 1053
Thr Thr Tyr Ser Ser Gly Asn Gln Asn Glu Lys Gln Ile
    340                 345                 350

GTG ACG ACT GAC TTG CGG CAG AAG TGC ACG GAG TCT CAC 1092
Val Thr Thr Asp Leu Arg Gln Lys Cys Thr Glu Ser His
            355                 360

ACG GGC ACC TCA GCC TCT GCC CCC TTA GCA GCC GGC ATC 1131
Thr Gly Thr Ser Ala Ser Ala Pro Leu Ala Ala Gly Ile
365                 370                 375

ATT GCT CTC ACC CTG GAG GCC AAT AAG AAC CTC ACA TGG 1170
Ile Ala Leu Thr Leu Glu Ala Asn Lys Asn Leu Thr Trp
        380                 385                 390

CGG GAC ATG CAA CAC CTG GTG GTA CAG ACC TCG AAG CCA 1209
Arg Asp Met Gln His Leu Val Val Gln Thr Ser Lys Pro
                395                 400

GCC CAC CTC AAT GCC AAC GAC TGG GCC ACC AAT GGT GTG 1248
Ala His Leu Asn Ala Asn Asp Trp Ala Thr Asn Gly Val
    405                 410                 415

GGG CGG AAA GTG AGC CAC TCA TAT GGC TAC GGG CTT TTC 1287
Gly Arg Lys Val Ser His Ser Tyr Gly Tyr Gly Leu Leu
            420                 425

GAC GCA GGC GCC ATG GTG GCC CTG GCC CAG AAT TGG ACC 1326
Asp Ala Gly Ala Met Val Ala Leu Ala Gln Asn Trp Thr
430                 435                 440

ACA GTC GCC CCC CAG CGG AAG TGC ATC ATC GAC ATC CTC 1365
Thr Val Ala Pro Gln Arg Lys Cys Ile Ile Asp Ile Leu
        445                 450                 455

ACC GAG CCC AAA GAC ATC GGG AAA CGG CTC GAC CTC CGG 1404
Thr Glu Pro Lys Asp Ile Gly Lys Arg Leu Glu Val Arg
                460                 465

AAC ACC GTG ACC GCG TCC CTG GGC GAG CCC AAC CAC ATC 1443
Lys Thr Val Thr Ala Cys Leu Gly Glu Pro Asn His Ile
    470                 475                 480
```

```
ACT CGG CTG GAG CAC GCT CAG GCG CGG CTC ACC CTG TCC 1482
Thr Arg Leu Glu His Ala Gln Ala Arg Leu Thr Leu Ser
            485                 490

TAT AAT CGC CGT GGC GAC CTG GCC ATC CAC CTG GTC AGC 1521
Tyr Asn Arg Arg Gly Asp Leu Ala Ile His Leu Val Ser
495                 500                 505

CCC ATG GGC ACC CGC TCC ACC CTG CTG GCA GCC AGG CCA 1560
Pro Met Gly Thr Arg Ser Thr Leu Leu Ala Ala Arg Pro
        510                 515                 520

CAT GAC TAC TCC GCA GAT GGG TTT AAT GAC TGG GCC TTC 1599
His Asp Tyr Ser Ala Asp Gly Phe Asn Asp Trp Ala Phe
                525                 530

ATG ACA ACT CAT TCC TGG GAT GAC GAT CCC TCT GGG GAG 1638
Met Thr Thr His Ser Trp Asp Glu Asp Pro Ser Gly Glu
    535                 540                 545

TGG GTC CTA GAG ATT GAA AAC ACC AGC GAA GCC AAC AAC 1677
Trp Val Leu Glu Ile Glu Asn Thr Ser Glu Ala Asn Asn
            550                 555

TAT GGG ACG CTG ACC AAC TCC ACC CTC GTA CTC TAT GGC 1716
Tyr Gly Thr Leu Thr Lys Phe Thr Leu Val Leu Tyr Gly
560                 565                 570

ACC GCC CCT GAC GGG CTC CCC GTA CCT CCA GAA AGC AGT 1755
Thr Ala Pro Glu Gly Leu Pro Val Pro Pro Glu Ser Ser
        575                 580                 585

GGC TGC AAG ACC CTC ACG TCC AGT CAG GCC TGT GTG GTG 1794
Gly Cys Lys Thr Leu Thr Ser Ser Gln Ala Cys Val Val
                590                 595

TGC GAG GAA GGC TTC TCC CTC CAC CAG AAG AGC TGT GTC 1833
Cys Glu Glu Gly Phe Ser Leu His Gln Lys Ser Cys Val
    600                 605                 610

CAG CAC TGC CCT CCA GGC TTC GCC CCC CAA GTC CTC GAT 1872
Gln Asn Cys Pro Pro Gly Phe Ala Pro Gln Val Leu Asp
            615                 620

ACG CAC TAT AGC ACC GAG AAT GAC GTG GAG ACC ATC CGG 1911
Thr Asn Tyr Ser Thr Glu Asn Asp Val Glu Thr Ile Arg
625                 630                 635

GCC AGC GTC TGC GCC CCC TGC CAC GCC TCA TGT GCC ACA 1950
Ala Ser Val Cys Ala Pro Cys His Ala Ser Cys Ala Thr
        640                 645                 650
```

```
TGC CAG GGG CCG GCC CTG ACA GAC TGC CTC AGC TGC CCC 1989
Cys Gln Gly Pro Ala Leu Thr Asp Cys Leu Ser Cys Pro
                655                 660

AGC CAC GCC TCC TTG GAC CCT GTG GAG CAG ACT TGC TCC 2028
Ser His Ala Ser Leu Asp Pro Val Glu Gln Thr Cys Ser
665                 670                 675

CGG CAA AGC CAG AGC AGC CGA GAG TCC CCG CCA CAG CAG 2067
Arg Gln Ser Gln Ser Ser Arg Glu Ser Pro Pro Gln Gln
            680                 685

CAG CCA CCT CGG CTG CCC CCG GAG GTG GAG GCG GGG CAA 2106
Gln Pro Pro Arg Leu Pro Pro Glu Val Glu Ala Gly Gln
690                 695                 700

CGG CTG CGG GCA GGG CTG CTG CCC TCA CAC CTG CCT GAG 2145
Arg Leu Arg Ala Gly Leu Leu Pro Ser His Leu Pro Glu
        705                 710                 715

GTG GTG GCC GGC CTC AGC TGC GCC TTC ATC GTG CTG GTC 2184
Val Val Ala Gly Leu Ser Cys Ala Phe Ile Val Leu Val
                720                 725

TTC GTC ACT GTC TTC CTG CTC CTG CAG CTG CGC TCT GGC 2223
Phe Val Thr Val Phe Leu Val Leu Gln Leu Arg Ser Gly
    730                 735                 740

TTT AGT TTT CGG GGG GTG AAG GTG TAC ACC ATG GAC CGT 2262
Phe Ser Phe Arg Gly Val Lys Val Tyr Thr Met Asp Arg
            745                 750

GGC CTC ATC TCC TAC AAG GGG CTG CCC CCT GAA GCC TGG 2301
Gly Leu Ile Ser Tyr Lys Gly Leu Pro Pro Glu Ala Trp
755                 760                 765

CAG GAG GAG TGC CCG TCT GAC TCA GAA GAG GAC GAG GGC 2340
Gln Glu Glu Cys Pro Ser Asp Ser Glu Glu Asp Glu Gly
        770                 775                 780

CGG GGC GAG AGG ACC GCC TTT ATC AAA GAC CAG AGC GCC 2379
Arg Gly Glu Arg Thr Ala Phe Ile Lys Asp Gln Ser Ala
                785                 790

CTC TGA                                             2385
Leu End
    795
```

It is presently and theoretically contemplated that the specific mechanism underlying enhanced expression of function γ-carboxylated proteins resides in the expression of DNA encoding PACE in mammalian cells which increases the efficiency of γ-carboxylation, a post-translational modification required for biological activity of certain mature proteins. The method is especially useful in the processing of vitamin K-dependent blood coagulation proteins. More specifically the method is useful in processing and γ-carboxylating other proteins including Protein C, Protein S, Prothrombin Factor IX, Factor VII, Factor X and bone γ-carboxyglutamate protein. For example, co-expression with PACE (SEQ ID NO: 2) with such a propeptide permits high level recombinant expression of biologically active mature proteins.

In addition, high levels of recombinant expression of functional proteins can also be achieved by use of the present method by expressing PACE (SEQ ID NO: 2) with more completely processed proteins expressed from other genes. For example, coexpression of PACE (SEQ ID NO: 2) with non-Vitamin K dependent propeptides which require cleavage but not γ-carboxylation for biological activity may produce high yields of functional mature proteins.

One such protein which may be expressed in high functional yields by the present method is bone morphogenic protein (BMP), particularly BMP-2 [see, e.g., E. Wang et al, Proc. Natl. Acad. Sci. USA, 87:2220-2224 (1990), which is incorporated by reference herein for information about that protein]. Other such proteins which may be produced in high functional yields by the present invention include tumor growth factor β (TGF-β) and platelet-derived growth factor (PDGF), among others.

Further, the present invention also encompasses the use of recombinant-derived PACE (SEQ ID NO: 2) for in vitro processing of nerve growth factor and monobasic propiomelanocortin. PACE (SEQ ID NO: 2) may also be useful in the processing of proteins, such as insulin, and for the maturation of viruses, such as HIV and Hepatitis C, which also require precursor processing at paired basic amino acid residues.

Transfection of a DNA sequence encoding PACE (SEQ ID NO: 1) and a DNA sequence for a selected propeptide precursor into a mammalian cell can be effected via one or more recombinant vectors carrying PACE (SEQ ID NO: 1), the mammalian propeptide, or both, using materials and methods conventional in heterologous gene expression in mammalian cells.

Host cells transformed with the one or more vectors carrying the PACE DNA (SEQ ID NO: 1) and the selected precursor DNA are selected, e.g. by conventional means, and may then be cultured under suitable conditions if desired, with amplification of one or both introduced genes. The method of this present invention therefore comprises culturing a suitable cell or cell line, which has been transformed with a DNA sequence coding for PACE (SEQ ID NO: 1) and a DNA sequence coding for the selected precursor, each coding sequence under the control of a transcriptional regulatory sequence. The expressed mature protein is then recovered, isolated and purified from the culture medium (or from the cell, if expressed intracellularly) by appropriate means known to one of skill in the art.

Suitable cells or cell lines for this method are mammalian cells, such as Chinese hamster ovary cells (CHO), the monkey COS-1 cell line or murine 3T3 cells derived from Swiss, Balb-c or NIH mice. The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening and product production and purification are known in the art.

See, e.g., Gething and Sambrook, Nature, 293:620-625 (1981), or alternatively, Kaufman et al, Mol. Cell. Biol., 5(7):1750-1759 (1985) or Howley et al, U. S. Patent 4,419,446. Another suitable mammalian cell line is the CV-1 cell line.

Further exemplary mammalian host cells include particularly primate cell lines and rodent cell lines, including transformed cell lines. Normal diploid cells, cell strains derived from in vitro culture of primary tissue, as well as primary explants, are also suitable. Candidate cells may be genotypically deficient in the selection gene, or may contain a dominantly acting selection gene. Other suitable mammalian cell lines include but are not limited to, HeLa, human adenovirus transformed 293 cells, mouse L-929 cells, BHK or HaK hamster cell lines.

The present invention also provides recombinant DNA molecules, or vectors, for use in the method of expression of active mature proteins, such as those described above. A single vector can carry the PACE DNA (SEQ ID NO: 1) and another vector can carry the selected precursor DNA, each under the control of a selected expression control sequence. Alternatively, both the PACE (SEQ ID NO: 1) and precursor DNA sequences may be carried on a single recombinant vector molecule in which

case they may be operably linked to respective expression control sequences or may share a common expression control sequence. In general, the vectors employed will contain selected regulatory sequences in operative association with the DNA coding sequences of PACE (SEQ ID NO: 1) and selected precursor and capable of directing the replication and expression thereof in selected host cells.

The vector used in the examples below is pMT3, a derivative of the previously described vector pMT2 [R. Kaufman, Mol. Cell. Biol., 9:946-958 (1989)]. The mammalian cell expression vectors described herein may be synthesized by techniques well known to those skilled in this art. The components of the vectors, e.g. replicons, selection genes, enhancers, promoters, and the like, may be obtained from natural sources or synthesized by known procedures. [See, Kaufman et al, J. Mol. Biol., 159:511-521 (1982); and Kaufman, Proc. Natl. Acad. Sci., USA, 82:689-693 (1985)].

Alternatively, the vector DNA may include all or part of the bovine papilloma virus genome [Lusky et al, Cell, 36:391-401 (1984)] and be carried in cell lines such as C127 mouse cells as a stable episomal element.

The transformation of these vectors into appropriate host cells can result in expression of the selected mature proteins. Other appropriate expression vectors of which numerous types are known in the art for mammalian expression can also be used for this purpose.

The following examples illustratively describe the construction of plasmids for the expression and production of PACE (SEQ ID NO: 2) in mammalian cells, and the co-expression of PACE (SEQ ID NO: 2) and the blood coagulation factor, Factor IX, in mammalian cells. These examples are for illustration and do not limit the scope of the present invention.

Example 1 - Plasmid Construction and Expression of PACE cDNA in COS-1 Cells

A 2.47 kbp PACE cDNA fragment [Chiron Corporation, California] is employed, which includes the published 794-codon PACE coding sequence and 74 bases of 3'-untranslated sequence before a SalI linker [A.M.W. van den Ouweland et al, cited above] (SEQ ID NO: 1)/(SEQ ID NO: 2). At the 5'-end, the sequence immediately preceding the ATG was modified to conform to the consensus translation start site using a EcoRI oligonucleotide adapter.

The 2.47 kbp (EcoRI-SalI) PACE cDNA fragment (SEQ ID NO: 1) was inserted into the SV40-based expression vector pMT3 to generate the plasmid pMT3-PACE. pMT3 has been deposited with the American Type Culture Collection (ATCC), Rockville, MD (USA) under Accession Number ATCC 40348. The pMT3 vector is a derivative of the previously described vector pMT2 [Kaufman, cited above] starting with pMT2-vWF, which is deposited at the American Type Culture Collection, Rockville, MD (USA), Accession Number ATCC #67122; see PCT application PCT/US87/00033]. To form pMT3, the DHFR coding region on the 3' side of the cloning site in pMT2 is removed.

One skilled in the art can also construct other mammalian expression vectors comparable to the pMT3/PACE vector by, e.g. inserting the DNA sequence of PACE (SEQ ID NO: 1) from pMT3 into another vector, such as pJL3, pJL4 [Gough et al., EMBO J., 4:645-653 (1985)], employing well-known recombinant genetic engineering techniques.

pMT3-PACE was purified and introduced for transient expression into SV40-transformed monkey kidney cells (COS-1) using a calcium phosphate transfection protocol as described in Chen, C. A., and Okayama, H., BioTechniques, 6:632-638 (1988). Cells were transfected with 40 μg of plasmid or, in the case of co-transfections, an equimolar ratio of plasmids totalling 60 μg per 10 cm dish in 10 ml of medium.

Mammalian host cells other than COS cells may also be employed in PACE (SEQ ID NO: 2) expression. For example, preferably for stable integration of the vector DNA, and for subsequent amplification of the integrated vector DNA, both by conventional methods, CHO may be employed as a mammalian host cell of choice.

To monitor PACE (SEQ ID NO: 2) synthesis, COS-1 cell products were radiolabeled 48-60 hours following transfection using ^{35}S-Met and ^{35}S-Cys in medium lacking Cys and Met. Cells were lysed in NP-40 lysis buffer after a 30 minute pulse period or were chased by removing the labeling medium and replacing it with complete medium for additional incubation. Cell extracts and conditioned medium were treated with protease inhibitors and immunoprecipitated as described in Wise et al, Cell, 52:229-236 (1988).

Immunoprecipitates were performed with rabbit anti-PACE antiserum produced against a PACE-E. coli fusion protein. Rabbit anti-PACE antiserum was generated against the catalytic domain of PACE by expression of amino acids 146 to 372 of PACE (SEQ ID NO: 2) as a human superoxide dismutase (SOD) fusion protein in E. coli. The DNA fragment for expression was generated by polymerase chain reaction (PCR) and cloned into the superoxide dismutase (SOD) fusion vector pTAC7 [Steimer et al, J. Virol., 58:9 (1986)].

The induced fusion protein was purified by preparative polyacrylamide gel electrophoresis, eluted and used to immunize rabbits in complete Freunds adjuvant. The immunoprecipitated samples were then analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The gels were prepared for fluorography in EnHance (Dupont).

In control lysates from COS-1 cells that did not receive pMT3-PACE, no immunoreactive proteins were detected. However, in extracts from pMT3-PACE transfected cells, immunoreactive species were detected that migrated primarily as a doublet of approximately 90 kD. Treatment of these PACE immunoprecipitates with the endoglycosidase enzyme, N-Glycanase, shifted the electrophoretic mobility of the PACE (SEQ ID NO: 2) consistent with the presence of asparagine-linked oligosaccharides.

Secreted products were analyzed from conditioned medium following a 12 hour chase period in medium containing an excess of unlabeled amino acids. Immunoprecipitations of the conditioned medium from pMT3-PACE transfected cells detected an immunoreactive protein migrating at 75 kD. The relative quantity of the 75kD PACE protein observed in the conditioned medium was 5 to 10 fold less than that remaining inside the cell at the

12 hour chase period. This secreted species may represent a truncated molecule missing the transmembrane domain, possibly the result of auto-proteolysis at the paired arginine residues, 497-498, due to the large overproduction of PACE (SEQ ID NO: 2) in the transfected COS-1 cells.

More extensive pulse-chase experiments demonstrated that the PACE translation product does not accumulate to high levels inside the cell compared to another integral membrane glycoprotein (influenza hemagglutinin) when synthesized at similar levels.

Example 2 - Co-Expression of PACE and Factor IX

A CHO cell line producing recombinant Factor IX (IC4) [the IC4 cell line is described in Kaufman et al, J. Biol. Chem., 261:9622-9628 (1986)] and Factor IX sequences were transfected with the PACE cDNA (SEQ ID NO: 1) described above in Example 1 operatively linked to another amplifiable marker, adenosine deaminase. The vector MT3SV2Ada [R.J. Kaufman et al, Meth. Enzym., 185:537-566 (1990)] was chosen for PACE expression because it contains a selectable ADA transcription unit but no DHFR sequences and the PACE fragment could easily be inserted after digestion of the vector with EcoR1 and Sal1.

A vector fragment was isolated from low melt agarose, ligated in a ratio of 5:1 (fragment to vector), diluted in 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and used to transform DH5 bacteria [Dr. Douglas Hanahan, Cold Spring Harbor, New York]. A nick-translated, ^{32}P labelled PACE fragment was prepared and used for filter hybridization to screen transformed colonies.

Positively hybridizing colonies were isolated and DNA prepared for digestion with EcoR1 and Sal1 for confirmation of PACE (SEQ ID NO: 1) insertion and with Bgl II for correct orientation of the fragment with respect to adenovirus major late promoter in the vector.

DNA from one colony was isolated for electroporation into the Factor IX producing cells, IC4. Pools of colonies have been selected for amplification by growth in 1.0 μM 2'-deoxycoformycin (DCF). The presence of PACE (SEQ ID NO: 2) in these amplified lines was confirmed by ^{35}S-methionine labelling and immunoprecipitation.

Biological activity of the Factor IX protein in the PACE/IX pools was analyzed by clotting assay, performed as described in Kaufman et al, J. Biol. Chem., 261:9622-9628 (1986). Cells were plated in p60 tissue culture dishes. The next day medium was reduced (1.5 ml) and changed to α "defined" + 1 μg/ml Vitamin K3.

The PACE/Factor IX pools were found to secrete between 2.0 and 3.1 fold more Factor IX biological activity than the original IC4 cell line. The results of a radioimmunoassay indicated increased levels of γ-carboxylated protein. These results are illustrated in Table I below.

TABLE I

Factor IX assays in original IC4 and PACE Co-expressing Cell Lines

Cell	CLOTTING ASSAY U/ml (pg/cell)	CLOTTING ASSAY U/ml (pg/cell)	RIA GLA µg/mL	RIA TOTAL µg/mL (pg/cell)	GLA/TOTAL
IC4	.28 (.32)	.18 (.18)	.1	20 (30)	.5%
Co-expressors	0.1 µM DCF	1.0 µM DCF	5 µM DCF		
A	.72 (.89) 2.7x	.45 (.48) 2.6x	.69	20 (29)	3.4%
B	.53 (.76) 2.3x	.39 (.41) 2.3x	1.05	22 (27)	4.8%
C	.66 (.73) 2.2x	.35 (.41) 2.3x	.17	19 (54)	.8%
D	.46 (.66) 2.0x	.55 (.55) 3.1x	1.14	17 (24)	6.7%
E	.67 (.80) 2.5x	.49 (.52) 2.9x	.3	11 (34)	2.7%

From the first electroporation of MT3 PACE Ada into IC4 cells, cells were selected in α medium with 10% dialyzed fetal calf serum, penicillin, streptomycin, glutamine, 200 μM Methotrexate and Adenosine, alanosine, uridine and 0.1μM DCF. Approximately 25 colonies were observed in plates that did not receive DNA.

A second electroporation performed was selected in the same manner and approximately 100 colonies were pooled into each of the 5 pools. Again, no colonies were observed on plates that did not receive DNA.

Expression of PACE (SEQ ID NO: 2) was detected in each pool by 30 minute pulse with ^{35}S Methionine followed by 2 hour chase and immunoprecipitation of cell extracts with α PACE antibody [Chiron Corporation, California]. In cells which express higher levels of PACE (SEQ ID NO: 2) as a result of selection for further DCF resistance, secretion up to 10-fold greater levels of γ-carboxylated Factor IX was observed compared to the original IC4 cell line.

The coexpression of PACE (SEQ ID NO: 2) did not produce any detectable change in the size of the Factor IX protein as monitored by immunoprecipitation with α FIX antibody [Hybridtech] and SDS gel electrophoresis.

Numerous modifications and variations of the present invention are included in the above-identified specification and are expected to be obvious to one of skill in the art. Such modifications and alterations to the compositions and processes of the present invention are believed to be encompassed in the scope of the claims appended hereto.